Karlsruher Institut für Technologie

12. Übung – TGI

(mit Teil 2 der 11. Übung)

Lorenz Hübschle-Schneider, Tobias Maier

INSTITUT FÜR THEORETISCHE INFORMATIK, PROF. SANDERS

Gedächtnislose Quellen

- Gibt Zeichen aus einem Alphabet $\Sigma = \{a_1, \ldots, a_n\}$ aus
- Zeichen a_i wird mit **fixer** Wahrscheinlichkeit p_i erzeugt, $\sum_{i=1}^{n} p_i = 1$
- Zeichen werden **unabhängig** erzeugt: Die Wahrscheinlichkeit, a_i zu bekommen, hängt **nicht** von den vorher erzeugten Zeichen ab

Gedächtnislose Quellen

- Gibt Zeichen aus einem Alphabet $\Sigma = \{a_1, \ldots, a_n\}$ aus
- Zeichen a_i wird mit **fixer** Wahrscheinlichkeit p_i erzeugt, $\sum_{i=1}^{n} p_i = 1$
- Zeichen werden **unabhängig** erzeugt: Die Wahrscheinlichkeit, a_i zu bekommen, hängt **nicht** von den vorher erzeugten Zeichen ab

Aus Shannons berühmtem Quellencodierungstheorem folgt:

Die bestmögliche Kompression der Ausgabe einer gedächtnislosen Quelle wird erreicht, wenn Zeichen a_i mit $\log_2(1/p_i)$ Bits codiert wird.

Gedächtnislose Quellen

- Gibt Zeichen aus einem Alphabet $\Sigma = \{a_1, \ldots, a_n\}$ aus
- Zeichen a_i wird mit **fixer** Wahrscheinlichkeit p_i erzeugt, $\sum_{i=1}^{n} p_i = 1$
- Zeichen werden **unabhängig** erzeugt: Die Wahrscheinlichkeit, a_i zu bekommen, hängt **nicht** von den vorher erzeugten Zeichen ab

Aus Shannons berühmtem Quellencodierungstheorem folgt:

Die bestmögliche Kompression der Ausgabe einer gedächtnislosen Quelle wird erreicht, wenn Zeichen a_i mit $\log_2(1/p_i)$ Bits codiert wird.

Darum können wir nicht hoffen, die Ausgabe einer gedächtnislosen Quelle S im Schnitt mit weniger Bits pro Zeichen zu codieren als

$$\mathrm{H}(S) = \sum_{i=1}^{n} p_i \log_2 (1/p_i) = -\sum_{i=1}^{n} p_i \log_2 p_i$$

Gedächtnislose Quellen

- Gibt Zeichen aus einem Alphabet $\Sigma = \{a_1, \ldots, a_n\}$ aus
- Zeichen a_i wird mit **fixer** Wahrscheinlichkeit p_i erzeugt, $\sum_{i=1}^{n} p_i = 1$
- Zeichen werden **unabhängig** erzeugt: Die Wahrscheinlichkeit, a_i zu bekommen, hängt **nicht** von den vorher erzeugten Zeichen ab

Aus Shannons berühmtem Quellencodierungstheorem folgt:

Die bestmögliche Kompression der Ausgabe einer gedächtnislosen Quelle wird erreicht, wenn Zeichen a_i mit $\log_2(1/p_i)$ Bits codiert wird.

Darum können wir nicht hoffen, die Ausgabe einer gedächtnislosen Quelle S im Schnitt mit weniger Bits pro Zeichen zu codieren als

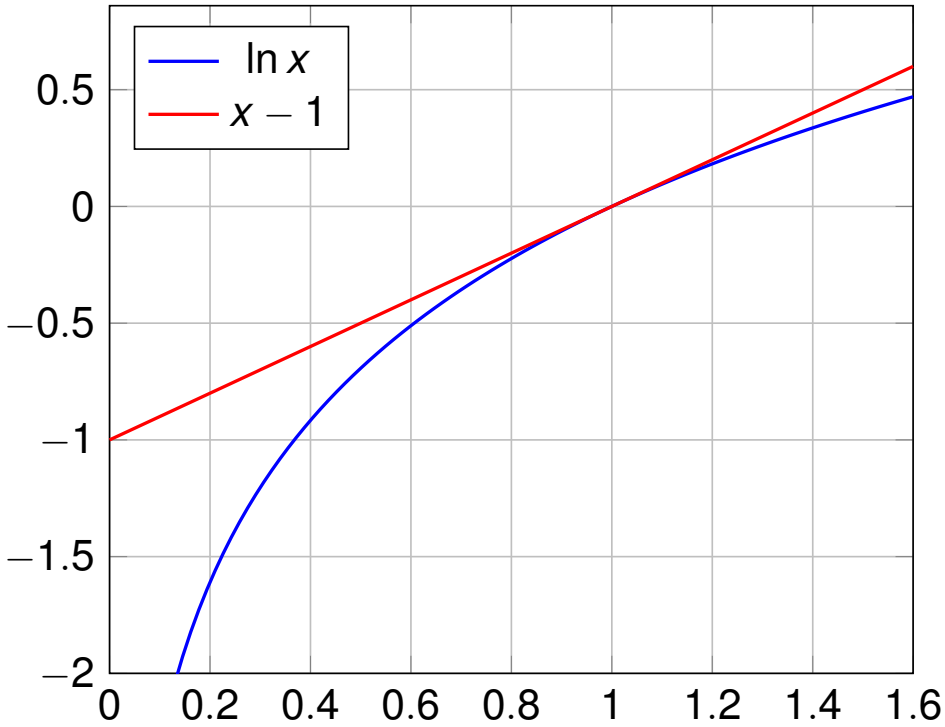
$$\mathrm{H}(S) = \sum_{i=1}^{n} p_i \log_2 (1/p_i) = -\sum_{i=1}^{n} p_i \log_2 p_i$$

$\mathrm{H}(S)$ heißt auch die *Entropie* der Quelle.

Entropiemaximierung

Entropie ist maximal, wenn $p_1 = \ldots = p_n = 1/n$ (alle gleich häufig).
Dann ist die Entropie $H_0 = \log_2 n$ bit

Beweis: Seien P, Q gedächtnislose Quellen mit je n Zeichen, Wahrscheinlichkeiten p_i bzw q_i. Abschätzung: $\ln x \leq x - 1$



Entropiemaximierung

Entropie ist maximal, wenn $p_1 = \ldots = p_n = 1/n$ (alle gleich häufig).
Dann ist die Entropie $H_0 = \log_2 n$ bit

Beweis: Seien P, Q gedächtnislose Quellen mit je n Zeichen, Wahrscheinlichkeiten p_i bzw q_i. Abschätzung: $\ln x \leq x - 1$

$\ln q_i - \ln p_i = \ln \frac{q_i}{p_i} \leq \frac{q_1}{p_1} - 1$. Multipliziere mit p_i und summiere über i:

$$\sum_{i=1}^{n} p_i \cdot (\ln q_i - \ln p_i) \leq \sum_{i=1}^{n} p_i \cdot \left(\frac{q_i}{p_i} - 1 \right)$$

Entropiemaximierung

Entropie ist maximal, wenn $p_1 = \ldots = p_n = 1/n$ (alle gleich häufig).
Dann ist die Entropie $H_0 = \log_2 n$ bit

Beweis: Seien P, Q gedächtnislose Quellen mit je n Zeichen, Wahrscheinlichkeiten p_i bzw q_i. Abschätzung: $\ln x \leq x - 1$

$\ln q_i - \ln p_i = \ln \frac{q_i}{p_i} \leq \frac{q_1}{p_1} - 1$. Multipliziere mit p_i und summiere über i:

$$\sum_{i=1}^{n} p_i \cdot (\ln q_i - \ln p_i) \leq \sum_{i=1}^{n} p_i \cdot \left(\frac{q_i}{p_i} - 1 \right)$$

$$-\sum_{i=1}^{n} p_i \ln p_i + \sum_{i=1}^{n} p_i \ln q_i \leq \sum_{i=1}^{n} q_i - \sum_{i=1}^{n} q_i = 0$$

Entropiemaximierung

Entropie ist maximal, wenn $p_1 = \ldots = p_n = 1/n$ (alle gleich häufig).
Dann ist die Entropie $H_0 = \log_2 n$ bit

Beweis: Seien P, Q gedächtnislose Quellen mit je n Zeichen, Wahrscheinlichkeiten p_i bzw q_i. Abschätzung: $\ln x \leq x - 1$

$\ln q_i - \ln p_i = \ln \frac{q_i}{p_i} \leq \frac{q_1}{p_1} - 1$. Multipliziere mit p_i und summiere über i:

$$\sum_{i=1}^{n} p_i \cdot (\ln q_i - \ln p_i) \leq \sum_{i=1}^{n} p_i \cdot \left(\frac{q_i}{p_i} - 1 \right)$$

$$-\sum_{i=1}^{n} p_i \ln p_i + \sum_{i=1}^{n} p_i \ln q_i \leq \sum_{i=1}^{n} q_i - \sum_{i=1}^{n} q_i = 0$$

$$\Rightarrow -\sum_{i=1}^{n} p_i \ln p_i \leq -\sum_{i=1}^{n} p_i \ln q_i$$

Entropiemaximierung

Entropie ist maximal, wenn $p_1 = \ldots = p_n = 1/n$ (alle gleich häufig).
Dann ist die Entropie $H_0 = \log_2 n$ bit

Beweis: Seien P, Q gedächtnislose Quellen mit je n Zeichen, Wahrscheinlichkeiten p_i bzw q_i. Abschätzung: $\ln x \leq x - 1$

$$\sum_{i=1}^{n} p_i \ln p_i \;\leq\; -\sum_{i=1}^{n} p_i \ln q_i$$

Annahme: Alle Zeichen von Q sind gleich wahrscheinlich: $q_i = 1/n$

Entropiemaximierung

Entropie ist maximal, wenn $p_1 = \ldots = p_n = 1/n$ (alle gleich häufig).
Dann ist die Entropie $H_0 = \log_2 n$ bit

Beweis: Seien P, Q gedächtnislose Quellen mit je n Zeichen, Wahrscheinlichkeiten p_i bzw q_i. Abschätzung: $\ln x \leq x - 1$

$$\sum_{i=1}^{n} p_i \ln p_i \;\leq\; -\sum_{i=1}^{n} p_i \ln q_i$$

Annahme: Alle Zeichen von Q sind gleich wahrscheinlich: $q_i = 1/n$

$$\sum_{i=1}^{n} p_i \ln p_i \;\leq\; -\sum_{i=1}^{n} p_i \ln \frac{1}{n} \;=\; \ln n \cdot \sum_{i=1}^{n} p_i \;=\; \ln n$$

Entropiemaximierung

Entropie ist maximal, wenn $p_1 = \ldots = p_n = 1/n$ (alle gleich häufig).
Dann ist die Entropie $H_0 = \log_2 n$ bit

Beweis: Seien P, Q gedächtnislose Quellen mit je n Zeichen, Wahrscheinlichkeiten p_i bzw q_i. Abschätzung: $\ln x \leq x - 1$

$$\sum_{i=1}^{n} p_i \ln p_i \;\leq\; -\sum_{i=1}^{n} p_i \ln q_i$$

Annahme: Alle Zeichen von Q sind gleich wahrscheinlich: $q_i = 1/n$

$$\sum_{i=1}^{n} p_i \ln p_i \;\leq\; -\sum_{i=1}^{n} p_i \ln \frac{1}{n} \;=\; \ln n \cdot \sum_{i=1}^{n} p_i \;=\; \ln n$$

Dividiere durch $\ln 2$ und wir erhalten $\mathrm{H}(P) \leq \log_2 n$.
Der Maximalwert wird bei gleich wahrscheinlichen Zeichen erreicht. □

Decodierbarkeit

Gegeben ein Code C der Zeichen a_i mit ℓ_i Bits codiert. Ist C eindeutig decodierbar?

Beispiel: Codiere a mit **0**, b mit **10**, c mit **110** und d mit **101**.

Decodierbarkeit

Gegeben ein Code C der Zeichen a_i mit ℓ_i Bits codiert. Ist C eindeutig decodierbar?

Beispiel: Codiere a mit **0**, b mit **10**, c mit **110** und d mit **101**.

Dieser Code ist **nicht** eindeutig: **010110** kann abc oder adb sein.

Kraft-McMillan-Ungleichung

Für jeden eindeutig decodierbaren Binärcode gilt $\sum_{i=1}^{n} 2^{-\ell_i} \leq 1$

Decodierbarkeit

Gegeben ein Code C der Zeichen a_i mit ℓ_i Bits codiert. Ist C eindeutig decodierbar?

Beispiel: Codiere *a* mit **0**, *b* mit **10**, *c* mit **110** und *d* mit **101**.

Dieser Code ist **nicht** eindeutig: **010110** kann *abc* oder *adb* sein.

Kraft-McMillan-Ungleichung

Für jeden eindeutig decodierbaren Binärcode gilt $\sum_{i=1}^{n} 2^{-\ell_i} \leq 1$

Im Beispiel: $2^{-1} + 2^{-2} + 2^{-3} + 2^{-3} = 1$.
Die Kraft-McMillan-Ungleichung ist eine **Implikation**, keine Äquivalenz!

Beobachtung: Präfixfreie Codes sind immer eindeutig decodierbar.
Auch das ist eine Implikation und keine Äquivalenz.

Theoretische Grenzen der Kompression

Wie weit lässt sich ein Wort komprimieren?

- Entropie hat klare theoretische Grenzen (gedächtnislos)

Theoretische Grenzen der Kompression

Wie weit lässt sich ein Wort komprimieren?

- Entropie hat klare theoretische Grenzen (gedächtnislos)
- Betrachte zum Beispiel das Wort $w = ababab \ldots ab$ (n mal)
  $\Rightarrow$ Entropie jedes Zeichens ist 1

Theoretische Grenzen der Kompression

Wie weit lässt sich ein Wort komprimieren?

- Entropie hat klare theoretische Grenzen (gedächtnislos)
- Betrachte zum Beispiel das Wort $w = ababab\ldots ab$ (n mal)
  $\Rightarrow$ Entropie jedes Zeichens ist 1
- Intuitiv braucht w nur $\log n + k$ Speicher

Theoretische Grenzen der Kompression

Wie lässt sich eine Kompression Definieren?

- $(\langle \mathcal{M} \rangle, in)$ heißt Beschreibung von w
  $\Leftrightarrow$ w ist Ausgabe von TM $\mathcal{M}$ wenn sie auf in startet.
- Die kürzeste Beschreibung von w ($K(w)$) heißt Kolmogorov Komplexität.

$$K(w) = \min_{B \text{ beschreibt } w} |B|$$

Fakten zur Kolmogorov Komplexität

- Die Kolmogorov Komplexität ist unberechenbar!
  $\Leftarrow$ Probieren aller Beschreibungen scheitert am Halteproblem
- Triviale obere Schranke $K(w) \leq |w|$
- Es gibt unkomprimierbare Wörter jeder Länge n
  $(\exists_w |w| = n \wedge K(w) \geq n)$:
  - Annahme $\forall_w |w| = n \Rightarrow K(w) < n$
  - Abzählbarkeitsargument $|\Sigma|^n$ Wörter der Länge k oder kleiner, aber nur $|\Sigma|^{n-1}$ kürzere Beschreibungen

Einige Beispiele für obere Schranken

- $K((ab)^n) \leq \log(n) + k$
- $K(\text{erste } n \text{ Stellen von } \pi) \leq \log(n) + k$
- $K(a^{\approx 3.5 * 10^{18267}}) \leq |(\langle \text{busy beaver } 6 \rangle, \varepsilon)|$